

OK

5

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-235933

(43) 公開日 平成7年 (1995) 9月5日

(51) Int. Cl. <sup>6</sup> H04L 12/28 G06F 13/00 15/16	識別記号 351 M 7230-5B 7831-5K	庁内整理番号 F I H04L 11/00 G06F 15/16	技術表示箇所 310 D 400 Z
---	----------------------------------	---	--------------------------

審査請求 有 請求項の数5 (全 11 頁) 最終頁に続く

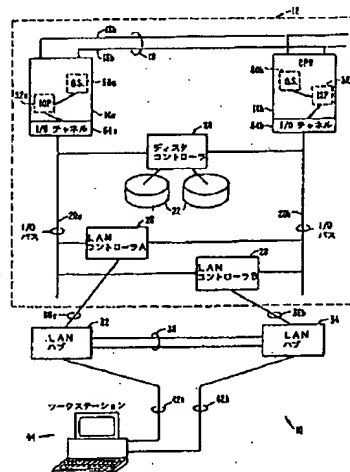
(21) 出願番号 特願平6-250637	(71) 出願人 391058071 タンデム コンピューターズ インコーポ レイテッド TANDEM COMPUTERS IN CORPORATED アメリカ合衆国 カリフォルニア州 95014 クーパティノ ノース タンタウ ア ベニュー 10435
(22) 出願日 平成6年 (1994) 10月17日	(72) 発明者 ケヴィン ジェイ ローウェット アメリカ合衆国 カリフォルニア州 95014 クーパティノ スティーブルチェス レーン 1134
(31) 優先権主張番号 08/137436	(74) 代理人 弁理士 中村 稔 (外6名)
(32) 優先日 1993年10月15日	
(33) 優先権主張国 米国 (U S)	

(54) 【発明の名称】 コンピュータ・システムのローカル・エリア・ネットワークへのフォールト・トレラント接続方法  
および装置

(57) 【要約】

【目的】 ネットワークのフォールト・トレラント接続を提供する。

【構成】 データ通信用入出力バスを備えたコンピュータ・システムが、一対のネットワーク制御装置によってネットワークに接続される。各ネットワーク制御装置は、一方で、一対のマルチ・ポート・ネットワーク・リピータ装置の対応する一つに接続される。これらのリピータ装置は、一方で、一対のネットワーク・リンクによって互いに接続される。少なくとも一つのワークステーションが、このリピータのそれぞれに接続される。ネットワーク制御装置の一つは、最初に、コンピュータ・システムからネットワークへのプライマリなデータ通信バスとして選択される。ネットワーク制御装置は、互いにメッセージを周期的に送信する。プライマリなネットワーク制御装置によるメッセージが送信されない場合には、プライマリな制御装置の選択が他方に切り換えられる。



## 【特許請求の範囲】

【請求項1】 データを通信するための入出力バス手段を備えた処理手段をネットワークにフォールト・トレラント接続するための方法であって、

第1および第2のネットワーク・コントローラ手段を前記入出力バス手段に接続し、

それぞれが、データを受信または送信するための複数のポートを有する一対のハブ手段であって、前記一対のハブ手段のそれぞれの前記複数のポートの一つが、対応する前記第1および第2のネットワーク・コントローラ手段とデータを送受信するために、対応する前記第1および第2のネットワーク・コントローラ手段と接続され、かつ、前記一対のハブ手段の一つの前記複数のポートのもう一つが、前記一対のハブ手段の他方の前記複数のポートの他の一つに接続されている、ハブ手段を提供し、少なくとも一つのステーション手段を前記一対のハブ手段のそれぞれの前記複数のポートの他の一つに接続し、および前記処理手段と前記一対のハブ手段との間のプライマリなデータ通信パスとして、前記第1または第2のネットワーク・コントローラ手段の一つを選択するステップ、を含む方法。

【請求項2】 前記第2および第1のネットワーク・コントローラ手段による受信のために、前記第1および第2のネットワーク・コントローラ手段からメッセージ・データをそれぞれ送信するステップ、を含む請求項1に記載の方法。

【請求項3】 前記第1または第2のコントローラ手段からのメッセージ・データが、前記第2または第1のコントローラ手段によりそれぞれ受信されないことが、前記プロセッサ手段に報告される、請求項1に記載の方法。

【請求項4】 データを通信するための入出力バス手段を備えた処理システムをネットワークにフォールト・トレラント接続するための装置であって、

前記入出力バス手段に接続される第1および第2のネットワーク・コントローラ手段、

それぞれが、データを受信または送信される複数のポートを有する一対の転送手段であって、前記一対の転送手段のそれぞれの前記複数のポートの一つが、対応する前記第1および第2のネットワーク・コントローラ手段とデータを送受信するために、対応する前記第1および第2のネットワーク・コントローラ手段と接続され、かつ、前記一対の転送手段の一つの前記複数のポートのもう一つが、前記一対の転送手段の他方の前記複数のポートの他の一つに接続されている、転送手段、

前記一対の転送手段のそれぞれの一つのポートに接続された少なくとも一つのステーション手段、および前記処理手段から前記一対の転送手段へのプライマリなデータ通信パスとして、前記第1または第2のネットワーク・

コントローラ手段の一つを選択する手段、を含む接続装置。

【請求項5】 前記少なくとも一つのステーション手段が、前記一対の転送手段のそれぞれの前記複数のポートの一つへの接続を、前記ステーション手段とのデータの通信のためのプライマリなパスとして選択する手段を含むものである、

請求項4に記載の接続装置。

## 【発明の詳細な説明】

10 【0001】

【産業上の利用分野】 この発明は、一般に、コンピュータ・システム・ネットワークの構成に関し、具体的には、発生する故障（フォールト）に対して許容性（トレラント）を有するネットワーク構成に関する。

【0002】

【従来の技術】 フォールト・トレラント・コンピュータ・システムは、以前から長く知られている。コンピュータ・タスクの高い利用性が望まれる環境においては、コンピュータ・システムは、様々なフォールト・トレラントの設計を備えている。フォールト・トレラント設計として、例えば、冗長なパーツ、素子またはモジュール、データ・チェック機能、およびスキームの補正等がある。フォールト・トレラント・コンピュータ・システムの一例は、米国特許第4,228,496号に記載されている。この特許は、例えば、それぞれが単一の故障箇所に制限する重複した装備の形態でのフォールト・トレラント機能を有するマルチプロセッサ・システムを教示している。

30 【0003】 このような高い利用性を有するコンピュータ・システムの一つの使用形態として、様々なネットワーク構成（例えば、ローカル・エリア・ネットワーク（LAN））における「サーバ」がある。このコンピュータ・システムは、そのネットワークに接続されたステーション（例えば、パーソナル・コンピュータ（PC）・システム）によって使用される様々なリソース（例えば、アプリケーション・プログラム、データ等）を記憶し、保持する。しかし、サーバ・ユニットがたとえ高い利用性を提供するように構成されていても、ネットワークが故障箇所を複数含むという点において、ネットワーク自体がその利用性を制限してしまう。例えば、ケーブルが切断され、または不注意に取り外されると、ネットワークの通信パスが除去され、おそらく一または複数のステーションがサーバから孤立するであろう。

【0004】 この問題の解決方法としては、十分余剰なネットワークを提供することとされてきた。しかし、この解決方法は、アプリケーション・プログラムを大量に再書き込みする必要がある。したがって、すぐに使用できるアプリケーション・プログラムの使用を事実上妨げることになる。

50 【0005】

【発明が解決しようとする課題】 このように、上述した米国特許第4,228,496号のようなマルチプロセッサ・システムによって提供されるフォールト・トレラントの高い利用性を、ネットワークに拡張する必要があることは明らかである。このフォールト・トレラント能力の拡張は、ワークステーション上またはネットワーク上で実行されているアプリケーションが、修正または変更される必要が全くないかほとんどないような方法で行われるべきである。

【0006】

【課題を解決するための手段】 この発明は、一般に、ネットワーク上で実行されるアプリケーションの変更を必要としない方法で、高い利用性を提供するフォールト・トレラント・ネットワークの構成に向けられている。この発明のより特徴的な点は、フォールト・トレラント・コンピュータ・システムのフォールト・トレラント結合をネットワーク構成に提供し、コンピュータ・システムのフォールト・トレラント能力をネットワークに拡張することである。

【0007】 この発明のネットワーク構成によると、コンピュータ・システムは、好ましくは、上述の米国特許第4,228,496号に教示されているようなものに基づいたマルチプロセッサ、およびこのシステムのプロセッサ・ユニットの少なくとも一つの入出力(I/O)バスに結合された一対のネットワーク・コントローラを備えている。一方、この一対のネットワーク・コントローラのそれぞれは、一対のマルチポート・ネットワーク・アクセス装置(好ましくは、コンセントレータ(集線装置)、すなわちIEEEスタンダード802.3の10BASET(または10BASE-T:IEEEの10BASE規格の一つ)配線の仕様を用いた星型配線のリピータ)の対応する一つに接続されている。他のステーションは、好ましくは2つのネットワーク・リンクによってネットワークに接続される。各リンクは、ステーションをネットワーク・アクセス装置の一つへ接続する。ネットワーク・アクセス装置自体は、好ましくは、通信できるように、一対のネットワーク・リンクによって互いに接続される。

【0008】 一対のネットワーク・コントローラが接続されたコンピュータ・システム(すなわち、サーバ)は、ネットワーク・コントローラの一つを、ネットワークへの「プライマリ」な通信バスとして選択し、もう一方のネットワーク・コントローラを「バックアップ」として指定する。したがって、処理システム(サーバ)とネットワークとの間のすべての通信は、このプライマリとして選択されたネットワーク・コントローラを介して行われる。

【0009】 2つのネットワーク・コントローラは、メッセージ・パケット(「ハートビート」メッセージと呼ばれる)を、ネットワークを介して周期的に互いに送信する。このメッセージ・パケットは、このパケットを送

信しているコントローラと接続パスとが、いまなお作動状態であることを示す信号である。他方からハートビート・メッセージを受信するネットワーク・コントローラの方に故障が発生すると、コンピュータ・システムは、そのコントローラを、「アンペアード」(unpaired)、すなわち使用不可能としてマークする。この使用不可能なネットワーク・コントローラが、たまたまプライマリとして選択されているネットワーク・コントローラであるならば、コンピュータ・システムは、最初にバックアップとして指定されているコントローラを、プライマリとして選択するように切り換え、最初にプライマリとして選択されたコントローラがアンペアードであることをログに記録する。

【0010】 したがって、これらのハートビート・メッセージは、2つの重要な機能をもつ。第1に、それらは、各LANコントローラ26および28がメッセージ・パケットを送信し受信できることを確立するために機能する。第2に、それらは、ネットワークが、LANコントローラ26および28が、一方がプライマリで、他方がバックアップのペアとして機能するように、構成されることを確立する。

【0011】 この発明によるネットワーク構成は、ある故障がネットワーク/ワークステーション上を実行されるアプリケーションに全体的に透明性を有することから、ネットワーク構成自体を再構成できる高い利用性を有するネットワーク・システムを提供する。

【0012】 この発明のこれらの特徴および他の特徴、ならびに利点は、この技術分野の専門家ならば、この発明についての以下の詳細な内容を、関連する図面を参照しながら読むことにより、明らかになるであろう。

【0013】

【実施例】 図1には、この発明の好ましい実施例として、ローカル・エリア・ネットワーク(LAN)の構成の概要が示されている。参照符号10で示されたこのネットワークには、このネットワークのネットワーク・サーバとして機能するコンピュータ・システム12が含まれている。このネットワークは、通信プロトコルとして、例えば、FDDI、トークン・リング等の他のプロトコルを使用することもできるが、ここでは、イーサネット・プロトコルを使用している。したがって、ネットワーク10で行われる通信は、従来からあるように、メッセージ・パケットを使用し、パケット発信ステーションの識別情報(ソース・アドレス)、受信ステーションの識別情報(デスティネーション・アドレス)、およびネットワークによって使用され、この特定のプロトコルにより提供される他の情報を含む。

【0014】 コンピュータ・システム12は、上述した米国特許第4,228,496号に示されているようにフォールト・トレラントのマルチプロセッサで構成されていることが好ましい。この技術分野の専門家ならば、コンピュー

タ・システム12に単一のプロセッサ・システムを使用できることを理解するであろう。しかし、この場合には、一箇所の故障の存在も認められず、このためにシステム全体のフォールト・トレラント能力が制限される。図に示すように、処理システム12は、プロセッサ・ユニット14 (14a、14b) を備えている。これらのプロセッサ・ユニットは、相互に接続され、バス16 (個別のバス16aおよび16bのペア) によってプロセッサ間通信を行う。また、各プロセッサ・ユニット14は、プロセッサ14aおよび14b用に、入出力 (I/O) バス20 (20a、20b) をそれぞれ備えている。

【0015】典型的には、特にサーバとして機能する場合には、処理システム12は、大容量記憶装置、例えばディスク・ユニット22を備えている。これらのディスク・ユニットは、デュアル・ポートのディスク・コントローラ24によって制御される。このディスク・コントローラ24は、一方で、I/Oバス20aおよび20bのそれぞれと接続されている。

【0016】また、デュアル・ポートのLANコントローラ26および28が、I/Oバス20aおよび20bに接続されている。一方で、このLANネットワーク・コントローラ26および28は、一対のLANアクセス装置32および34にそれぞれ接続され、コンピュータ・システム12をLANアクセス装置32および34に接続する。利用できるLANコントローラとして、ここでは、代表的に、カリフォルニアのキューパティノ (Cupertino) にあるタンデム・コンピュータ社によって製造された3615 LANコントローラがある。このようなLANコントローラは、マイクロプロセッサをベースとしたマシンという点で、「インテリジェント」装置と呼ばれるものである。

【0017】LANアクセス装置32および34は、それぞれ一対のコンセントレータ (集線装置)、すなわちIOBASET (またはIOBASE-T: IEEEのIOBASE規格の一つ) 配線を使用した星型配線リピータであることが好ましい。利用可能なアクセス装置として、カリフォルニアのサンタ・クララのアンジャーマン・バス社 (Ungermann-Bass Corporation) によって製造された2つのIOBASET コンセントレータ・カード (P/N ASM 32012) を含むACCESS/ONE (アンジャーマン・バス社の登録商標) のASC 2000キャビネットがある。このLANアクセス装置は、後述するように、「ハブ」 (hub) 32 および34と呼ばれる。この技術分野の専門家ならば明らかであろうが、他のタイプのメディア・アクセス装置 (例えば、ブリッジ) をこの発明の教示する範囲から逸脱することなく使用することもできる。

【0018】このように、LANコントローラ26は、リンク36aを介してLANハブ32の複数のポートの一つと接続され、LANコントローラ28は、リンク36bによってLANハブ34と接続されている。各LANハブ32および36の一対のポート (例えば、バックボーン・コネクシ

ョン・ポート) は、この二つのLANハブ32および34をリンク38によって互いに接続し、通信できるようにしている。また、LANハブ32および34のそれぞれ1つのポートは、リンク42aおよび42bによりワークステーション44に接続されている。

【0019】LANコントローラ26および28には、3つのメディア・アクセス制御 (MAC) アドレスがそれぞれ割り当てられている。第1のアドレスは、このコントローラを指定するためのユニークなアドレスである。このユニークなアドレスをデスティネーション・アドレスとして含んだメッセージ・パケットは、指定されたLANコントローラによってのみ受信される。第2のアドレスは、「ペア・ワイズ」 (pair-wise) アドレスであり、LANコントローラ26および28のペアを指定するものである。このようなペア・ワイズ・アドレスをデスティネーション・アドレスとして含むネットワークのメッセージ・パケットは、LANコントローラ26と28の両方で受信される。第3のアドレスは、グループMACアドレスである。このグループMACアドレスに設定されたメッセージ・パケットは、グループMACアドレスによって指定されるグループの全てのユニットまたはステーションによって受信される。LANコントローラ26および28がそのグループに割り当てられているならば、これらのコントローラにもこのパケットは受信される。

【0020】各プロセッサ・ユニット14は、オペレーティング・システム (OS) 50の制御の下に処理を行う。OS 50は、利用可能な関連する入出力プロセス (IOP) ・デバイス・ドライバ52を有し、このドライバにより、プロセッサ・ユニット14との通信が (プロセッサ・ユニット14とI/Oバス20とのインタフェースを形成するI/Oチャネル54を介して) 制御される。IOP 52は、LANコントローラ26および28に介して通信を制御する。

【0021】各LANコントローラは、LANハブ32、34およびネットワークと通信するのに必要なMACレイヤ・プロトコルを実行する。LANコントローラ26および28のペアは、グループMACアドレスに、いわゆる「ハートビート」と呼ばれるメッセージ・パケットを周期的に送信する。このパケットは、このペアとなっている他方のコントローラを含む、ネットワーク上の全ての他の装置によって受信される。したがって、例えば、LANコントローラ26は、その兄弟のLANコントローラ28に受信されるハートビート・メッセージ・パケットを周期的に送信する。このハートビート・メッセージ・パケットは、一方のLANコントローラが、その兄弟のLANコントローラに、この一方のLANコントローラがまだ処理状態にあり、かつ、このコントローラのネットワーク10 (および兄弟のLANネットワーク・コントローラ) への接続が存続していることを示すものである。

LANコントローラ26および28は、このようなハートビ

ート・メッセージ・パケットが受信されないと、その受信されないことをIOP52に報告する。IOP52は、それが制御するLANコントローラの状態のログ(経過記録)を、ソフトウェア状態マシン(図2)の形態で保持する。

【0022】ハートビート・メッセージ・パケットは、ソース・アドレスとして、そのパケットを送信したLANコントローラ26および28のユニークな識別アドレスを含む。受信したLANコントローラは、LAN上の他のトラフィックの報告だけでなく、その兄弟からのハートビート・メッセージを受信したことを周期的に報告する。ある特定のLANコントローラ(例えば、LANコントローラ28)がその兄弟(例えば、LANコントローラ26)からのメッセージを受信しなかったならば、このメッセージが受信されなかったことが、対応するIOP52に報告される。メッセージを送信しなかった(サイレントな)LANコントローラが、プライマリなものとして指定されていたならば、この対応するIOP52は、この指定を切り換える。バックアップとして形式的に指定されていたLANコントローラが、プライマリなものとなり、前者のプライマリな(現在はサイレントな)LANコントローラは、「アンペアード」(unpaired)として指定される。この「アンペアード」は、何らかの理由により、LANコントローラまたはそのネットワーク接続が適切に機能しなくなったことを示すものである。

【0023】2つのプロセッサ14は、一対のLANコントローラ26および28に接続されているので、これらのプロセッサは、いずれがこれらのLANコントローラを制御するのか、すなわちいずれが「プライマリ」なプロセッサ・ユニットになるのかを、2つのプロセッサ間で決定しなければならない。他のプロセッサは、バックアップになり、プライマリなものが機能しなくなったときのみ、処理を行うことになる(米国特許第4,228,496号参照)。

【0024】続いて、プライマリなプロセッサ・ユニット14のIOP52は、このプロセッサ・ユニット14とネットワーク10との間のプライマリな通信パスとして、LANコントローラ26の一つを(少なくとも最初に、大かれ少なかれ裁定的な方法で)選択する。したがって、プロセッサ・ユニット14とのすべての通信は、(IOP52の制御の下に)選択された「プライマリ」なLANコントローラ26を通じて行われる。一対のLANコントローラのもう一方(例えば、LANコントローラ28)は、バックアップとして機能する。このバックアップのLANコントローラが、ネットワークからのメッセージ・パケットを受信するかもしれないが、これらのパケットは、指定されたプロセッサ・ユニット14には送信されない。

【0025】一方がプライマリであり他方がバックアップ(すなわち、利用可能)であるというLANコントローラ26および28の状態は、IOP52によってソフトウェ

ア・状態マシンに保持される。IOPにより保持される状態マシンの状態図(状態遷移図)が、図2に示されている。この図は、5つの個別の状態を示し、これらの状態は、LANコントローラ26および28によって仮定される異なる状況(すなわち、各LANコントローラがプライマリ、バックアップ、および「アンペアード」によって示される利用不可能な状態)を示す。各状態は、次に示すものである。

- 10 【0026】1. コントローラ26がプライマリ、コントローラ28がバックアップ
2. コントローラ26がバックアップ、コントローラ28がプライマリ
3. コントローラ26がプライマリ、コントローラ28がアンペアード
4. コントローラ26がアンペアード、コントローラ28がプライマリ
5. 両方ともダウン

【0027】上述した用語「アンペアード」は、LANコントローラがバックアップとして利用できないことを示すものである。この利用できないことは、コントローラ自体が正しく作動しない、または例えばLANコントローラ26の場合にリンク36aのようなリンクが正しく接続されていないことに起因する。

【0028】プライマリとして選ばれたLANコントローラ26または28は、IOP52が送信すべき全てのメッセージ・パケット・トラフィックを通信のためにネットワーク10へ送信し、ネットワーク10からのメッセージ・トラフィックを受信するものとなる。このIOP52は、バックアップ(または明らかに、アンペアード)のLANコントローラを介してネットワークとは通信しない。

【0029】IOPに保持された状態マシンが状態変化を起こす6つの異なるイベントがある。これらのイベントは次の通りである。

【0030】1. ハートビート・メッセージが、LANコントローラ26によって送信され、LANコントローラ28によって受信されたものとして、IOP52に報告される(以下に示すテーブルでは、これは“HB msg 26 → 28”として示されている)。

2. ハートビート・メッセージが、LANコントローラ28によって送信され、LANコントローラ26によって受信されたものとして、IOP52に報告される(“HB msg 28 → 26”)。

3. LANコントローラ26が、LANコントローラ28からのハートビート・メッセージ・パケットではないMACメッセージ・パケット・トラフィックを受信したことを報告する(“26” rpts MAC”)。

4. LANコントローラ28が、LANコントローラ26からのハートビート・メッセージ・パケットではないMACメッセージ・パケット・トラフィックを受信したことを報告する(“28” rpts MAC”)。

5. LANコントローラ26が、サイレントであることを、すなわち、ハートビート・メッセージもMACメッセージ・パケット・トラフィックも受信されないことを報告する(“26” rpts silence)。

【0031】 6. LANコントローラ28が、サイレントであることを報告する(“28” rptssilence)。

【0032】 上記のイベント発生の結果として状態マシンで行われる3つの動作がある。これらの動作は、次のとおりである。

【0033】 1. 切り換える(すなわち、最初にバックアップであったLANコントローラをプライマリとして選択する)。

2. イベントの発生を無視する。

3. イベントの発生を報告する。

もちろん、行われる特定の動作は、IOPが保持する状態マシンの状態に依存する。以下に示す表1～表5は、発生する6つのイベントのそれぞれ対応して、行われる動作およびこのイベントと現在の状態に依存して状態マシンによって取られる新しい状態を示している。各表の左の列は、イベントである。中央の列は、イベントの発生により取られる動作である。各表の右の列は、動作の結果を示している(すなわち、同じ状態の留まるか、または新しい状態に変化することを示している)。

【0034】

【表1】

状態(1) コントローラ26 プライマリ; コントローラ28 バックアップ

イベント	動作	新たな状態
(1) HB msg 26 → 28	(2)無視	(1)-現状態維持
(2) HB msg 28 → 26	(2)無視	(1)-現状態維持
(3)"26" rpts MAC	(3)報告	(3)-26 プライマリ; 28 アンペアード
(4)"28" rpts MAC	(1)切り換え	(4)-26 アンペアード; 28 プライマリ
(5)"26" rpts silence	(1)切り換え	(4)-26 アンペアード; 28 プライマリ
(6)"28" rpts silence	(3)報告	(3)-26 プライマリ; 28 アンペアード

【0035】

【表2】

状態(2) コントローラ26 バックアップ; コントローラ28 プライマリ

イベント	動作	新たな状態
(1) HB msg 26 → 28	(2)無視	
(2) HB msg 28 → 26	(2)無視	
(3)"26" rpts MAC	(1)切り換え	(3)-26 プライマリ; 28 アンペアード
(4)"28" rpts MAC	(3)報告	(4)-26 アンペアード; 28 プライマリ
(5)"26" rpts silence	(3)報告	(4)-26 アンペアード; 28 プライマリ
(6)"28" rpts silence	(1)切り換え	(3)-26 プライマリ; 28 アンペアード

【0036】

【表3】

状態(3) コントローラ26 プライマリ; コントローラ28 アンペアード

イベント	動作	新たな状態
(1) HB msg 26 → 28	(3)報告	(1)-26 プライマリ; 28 バックアップ
(2) HB msg 28 → 26	(3)報告	(1)-26 プライマリ; 28 バックアップ
(3)"26" rpts MAC	(2)無視	(3)
(4)"28" rpts MAC	(2)無視	(3)
(5)"26" rpts silence	(3)報告	(5)-双方ダウン
(6)"28" rpts silence	(2)無視	(3)

【0037】

【表4】

状態(4) コントローラ26 アンペアード; コントローラ28 プライマリ

イベント	動作	新たな状態
(1) HB msg 26 → 28	(3)報告	(2)-26 バックアップ; 28 プライマリ

(2) HB msg 28 → 26	(3) 報告	(2)-26 バックアップ; 28 プライマリ
(3)"26" rpts MAC	(2) 無視	(4)
(4)"28" rpts MAC	(2) 無視	(4)
(5)"26" rpts silence	(2) 無視	(4)
(6)"28" rpts silence	(3) 報告	(5)-双方ダウン

【0038】

【表5】

状態(5) 両コントローラ ダウン

イベント	動作	新たな状態
(1) HB msg 26 → 28	(3) 報告	(1)-26 プライマリ; 28 バックアップ
(2) HB msg 28 → 26	(3) 報告	(2)-26 バックアップ; 28 プライマリ
(3)"26" rpts MAC	(3) 報告	(3)-26 プライマリ; 28 アンベアード
(4)"28" rpts MAC	(3) 報告	(4)-26 アンベアード; 28 プライマリ
(5)"26" rpts silence	(2) 無視	(5)
(6)"28" rpts silence	(2) 無視	(5)

【0039】例えば、表1は、この状態マシンが状態1にあるときに発生するイベントを示している。LANコントローラ26がプライマリとして選択され、LANコントローラ28はバックアップである。イベント1（LANコントローラ28がLANコントローラ26からハートビート・メッセージ・パケットを受信する）が発生すると、この動作は無視され（アクション2）、状態マシンは状態1に留まる。しかし、状態1において、イベント3（LANコントローラ26が、MACメッセージ・パケットのみを受信しハートビート・メッセージ・パケットを受信していないことをIOPに報告）が発生すると、このイベントは、LANコントローラ28がおそらく上記で概説した一つまたは複数の原因により利用できなくなっていることを示している。したがって、状態マシンは、状態1から状態3に移動し、LANコントローラ26は、プライマリを維持し、LANコントローラ28は、「アンベアード」として指定される。

【0040】IOPが保持する状態マシンが状態を維持するLANコントローラの処理のもう一つの例が、状態図（図2）の状態3を示す表3に示されている。この表3では、LANコントローラ26が処理システム12とLANハブ32および34との間のプライマリのデータ・パスの状態にあり、LANコントローラ28がもはやバックアップでなく、アンベアードの状態にあるときを示している。LANコントローラ28がLANコントローラ26からのハートビート・メッセージ・パケットの受信の報告を開始（イベント1）、またはLANコントローラ26がLANコントローラ28からのハートビート・メッセージの受信を報告する、という2つのイベントの一方または双方が発生したと仮定する。いずれか一方のイベントの発生は、LANコントローラ26およびそのネットワーク接続の双方またはいずれか一方が適切に機能していることを示す。表3は、これらのイベントの発生のいずれかに

よって行われる動作がそれらを制御するIOP52に報告することであることを示し、また、状態1へ状態変化することを示している（LANコントローラ26はネットワーク10へのプライマリな接続を維持し続け、LANコントローラ28の状態はアンベアードからバックアップに変化する）。表1～表5の残りのイベントについても、発生する上記6つのイベント、およびこれらのイベントの発生の結果として行われ、状態マシンのそのときの状態に依存して行われる動作を参照して、同様に解釈することができる。

【0041】IOPが保持する状態マシンの連続した状態追跡能力により、IOP52は、LANコントローラ26または28のいずれがネットワーク10へのプライマリな接続を行うものになるか、いずれが状態変化によりバックアップになるかを再構成することができることは注目すべきである。また、以下の説明で明らかになるが、これは、トランスペアレントな（透明性を有する）方法で、ネットワークの残りの部分のすべてについてなされる。例えば、LANコントローラ26をプライマリとして、LANコントローラ28をバックアップとして最初にそれぞれ選択した後に、LANコントローラ28がMACメッセージ・パケット・トラフィックのみの受信の報告を開始するか（すなわち、ハートビート・メッセージ・パケットがLANコントローラ28から受信されていない：イベント4）、またはLANコントローラ26がサイレントであることの報告する（イベント5）かのいずれかを、制御IOP52が確認したと仮定する。状態図（図2）および表1が示すように、これらのいずれのイベントの発生によっても、状態マシンは、状態1から状態4へ移行し、LANコントローラ26はアンベアードに指定され、LANコントローラ28はネットワーク10にプライマリ接続になる。IOP52は、ネットワーク10との以後のすべての通信が、LANコントローラ28を通じて行われること

を、そこに見るであろう。

【0042】ネットワーク接続におけるこの変化の透明性は、LANコントローラ26および28によって使用されるペア・ワイズ・アドレスによるものである。処理システム12に送信されるすべてのメッセージ・パケットは、このペア・ワイズ・アドレスを用いる。このアドレスは、上述したように、両方のLANコントローラで同じである。このように、ワークステーション44がメッセージ・パケットをコンピュータ・システム12に送信したときに、そのパケットのデスティネーション・アドレスは、LANコントローラ26および28の双方を指定している。そして、両方のコントローラが、このパケットを受信する。プライマリとして指定されているLANコントローラ26または28のみが、このメッセージ・パケットをそのプロセッサ・ユニット14（すなわち、プロセッサ14上で実行されているIOP52）に転送する。したがって、コンピュータ・システム12は、ネットワークの他のステーション（例えば、ワークステーション）の補助、または他のステーションの知識さえも用いることなく、ネットワーク10に最初にプライマリ接続されたもののどこかに発生した故障に対処するためにそれ自身を再構成することができる。

【0043】LANのハブ・ユニット32および34は、IEEEスタンダード802.1、特にこのスタンダードの極大木と学習ブリッジ・アルゴリズム（“Spanning Tree and Learnig Bridge Algorithm”）を実行するように設計されていることが好ましい。このアルゴリズムにより、LANハブ・ユニット32および34は、LANコントローラ26および28を含むネットワーク上のステーションへのトラフィックまたはステーションからのトラフィックに基づいて、ネットワークの構成およびステーションの配置を学習することができる。LANハブ・ユニット32および34は、LANコントローラ28へのメッセージ・パケット・トラフィックおよびLANコントローラ26からのメッセージ・パケット・トラフィックの認識を開始し、それ自身を再構成して、処理システム12のトラフィックをそのLANコントローラへ転送する。

【0044】また、この発明の好ましい実施例は、2つのリンク38を使用し、LANハブ32および34自身を相互接続する。IEEEスタンダード802.1に示される極大木と学習ブリッジ・アルゴリズムは、2つのリンク38の一つをプライマリとして、もう一つをバックアップとしてそれぞれ選択する。プライマリが故障すると、このアルゴリズムは、バックアップを選択する。

【0045】この発明に組み込まれたフォールト・トレラントの思想は、ワークステーション44を2つのLANハブ32および34のそれぞれに接続することにより、ワークステーション44をネットワーク10に接続する場合にも拡張される。これにより、たとえ一方が停止しても、もう一方の通信パスが提供される。ネットワーク10にワー

クステーション44を接続する2重リンクのためのインタフェースが、図3の参照符号60によって示されている。図に示すように、ネットワーク・インタフェース60は、マルチプレクサ（MUX）62および従来からあるLAN制御ロジック64を備えている。実際に実現する場合に、MUX62は、電気機械式リレーで実現される。このリレーは、LAN制御ロジック64に接続され、2つのネットワーク・リンク42のいずれか一方の選択処理を行う。MUX62は、ワークステーション44の中央処理装置（CPU）からの選択信号（SEL）の制御の下、処理を行う。選択されたリンク42は、ワークステーション44のネットワーク10の「プライマリ」な通信リンクになる。

【0046】LAN制御ロジック64は、ウルトラチップ（Ultrachip）・イーサネット・コントローラのような従来からあるネットワーク・コントローラ・チップであり、部品番号83C790QFで販売され、ニューヨークのハウパーズ（Haupauge）にあるスタンダード・マイクロシステムズ・コーポレーション（Standard Microsystems Corporation）により製造され、利用可能である。このようなネットワーク・コントローラ・デバイスは、一般には、LANのリンクをモニタするためのパルス検出回路を含み、リンクが故障していないことを調べるためにLANのリンクに取り付けられている。従来と同様に、LANハブ32および34は、リンク保全パルスを、メッセージ・パケット・トラフィックの存在しないときに、ポートのそれぞれから周期的に送信する。リンク保全パルスにより、リンク自身（およびリンクに接続されたポート）が良好な状態にあることが示される。このようなリンク保全パルスのいずれかが欠如すると、LAN制御ロジック64は、CPU68に欠如していることを報告する。

【0047】ネットワーク・インタフェース60をネットワーク・リンク42にポート接続する場合に、従来のAUI接続も使用できるが、IOBASETタイプの配線接続を用いることが好ましい。IOBASET配線接続は、リンク保全パルスの欠如をより早い時間で報告できるので好ましい。

【0048】最初に、CPU68は、リンク42の一つをプライマリなリンクとして選択し、LAN制御ロジック64にそのリンクを通知する。その後、ワークステーション44とネットワーク10との間の全てのトラフィックが、選択されたリンク42（たとえば、42a）のみで通信される。上述したように、LANハブ32、34からのリンク保全パルスの受信（または欠如）から決定されたプライマリ・リンクの保全に関する報告を、CPU68はLAN制御ロジック64から周期的に受信する。この報告が、プライマリなリンクを指定するものであるならば、CPU68は、MUX62を切り換え、形式的にはバックアップであったリンク42をLAN制御ロジック64に通知する。

【0049】最初にプライマリとして選択されなかったリンク（例えば、42b）は、バックアップ・リンクと呼



ばれる。周期的に(例えば、1/2秒ごとに)、CPUは、バックアップ・リンク42bを選択し、このリンクをLAN制御ロジック64に接続し、このリンクがLANハブ34(図1)によって送信されるリンク保全パルスを通信用しているかどうかを判定させる。リンク保全パルスの欠如が、CPU68に報告されないならば、CPU68は、何らかの理由により、リンク42bは利用できないものと決定する。

【0050】図1に戻り、LANコントローラ26、28が2つのプロセッサ・ユニット14に接続されていることに着目すると、各プロセッサ・ユニット14は、一対のLANコントローラ26および28に対して一つのIOP52を備えていることが分かる。一つのプロセッサ・ユニット14が追加のペアに接続されたならば、追加のIOP52が必要となるであろう。これは、図4に示されている。この図は、N個のプロセッサまで有する、より一般化されたマルチ・プロセッサ・システムとしての処理システム12を示している。米国特許第4,228,496号のマルチ・プロセッサ・システムでは、Nは好ましくは16に制限されている。

【0051】このように、図4に示すように、プロセッサ・ユニット14aおよび14bに加えて、プロセッサ・ユニット14nが存在し、追加の一対のLANコントローラ26'および28'が、I/Oとプロセッサ14b、14nのバス20b、20nとの間に接続されている。LANコントローラ26'および28'は、LANコントローラ26および28と同じ接続方法で、ネットワーク・リンク34a'および34b'によってLANハブ装置32および34にそれぞれ接続されている。

【0052】プロセッサ・ユニット14nは、(プロセッサ・ユニット14nのI/Oチャネル54nを介して)LANコントローラ26'および28'を操作するための入出力プロセス(IOP)・デバイス・ドライバ52nを備え、オペレーティング・システム(OS)50nの下で処理を行う。しかしながら、プロセッサ・ユニット14bは、2つの個別のIOPデバイス・ドライバを有する。一つのIOP52bは、LANコントローラ26および28を介してネットワーク10と入出力通信を制御するためのものであり、これとは別のIOPデバイス・ドライバ52b'は、第2のLANコントローラのペア26'および28'を介して通信を制御するためのものである。

【0053】この発明を、単一のワークステーション44について説明してきたが、この技術分野の専門家にとっては、実際には、多くの様々なタイプのステーションが存在することが明らかであろう。これらのステーションのそれぞれも、ワークステーション44と同じようにLANハブ32、34に接続され、すなわち、個別のリンク接続になることが好ましい。また、この発明は、図に示すように、処理システム14により(2つのプロセッサ・ユニット間に接続されたLANコントローラ26および28の各

ペアにより)実現されるが、コンピュータ・システム14は、マルチプロセッサ・システムによって提供されるフォールト・トレラント能力を避けることができ、単一ポートのLANコントローラがこのプロセッサをネットワークに接続する単一のI/Oパスのみを備えている。この単一プロセッサは、一箇所の故障を提供するが、一対のLANコントローラおよびネットワーク自身の構成によるネットワーク接続は、フォールト・トレラントなネットワーク設計を提供する。

10 【0054】さらに、この発明の好ましい実施例は、コンピュータ・システム12をネットワーク10に相互接続するのに、「インテリジェント」なLANコントローラを用いるが、場合によってはその必要は必ずしもない。この技術分野の専門家ならば、このようなインテリジェントなLANコントローラによって実行される処理の多くが、コンピュータ・システムの処理時間のコストを必要とするが、コンピュータ・システム上を実行されるIOPによって行われることを知っている。

【0055】特許請求の範囲に記載の発明に加えて以下  
20 の発明を開示する。

【0056】請求項2に記載の方法における送信ステップは、周期的に行われるものであってもよい。

【0057】また、請求項3に記載の方法は、前記第1および第2のネットワーク・コントローラ手段の最初に選択された一つからのメッセージ・データの受信の欠如が前記処理手段に報告されたときに、前記第1および第2のネットワーク・コントローラ手段のもう一つをプライマリなデータ通信パスとして選択するステップを含むことができる。

30 【0058】請求項1に記載の方法は、前記ハブ手段のそれぞれの前記複数のポートのさらに一つが、それらの間でデータを通信するために他方に接続され、前記ハブ手段の間のプライマリなデータ通信パスとして、ハブとハブとの間の接続の一つを選択するステップを含むことができる。

【0059】請求項4に記載の装置は、前記第1および第2のネットワーク・コントローラ手段が、少なくとも第2および第1のネットワーク・コントローラ手段によってそれぞれ受信されるメッセージ・データを送信する  
40 ための手段を含むことができる。

【0060】また、前記第1および第2のネットワーク・コントローラ手段は、メッセージ・データの受信が行われないことを前記処理手段に報告し、前記第1または第2のネットワーク・コントローラ手段の最初に選択された一つからのメッセージ・データは受信されないときには、前記処理手段は、第1および第2のネットワーク・コントローラ手段のもう一方をプライマリなデータ通信パスとして選択するものであってもよい。

【0061】請求項4に記載の装置において、第1および第2のネットワーク・コントローラ手段は、第2およ  
50

び第1のネットワーク・コントローラ手段にそれぞれ受信されるメッセージ・データを周期的に送信するものであり、プライマリな通信パスとして選択された第1または第2のネットワーク・コントローラ手段によるメッセージ・データを受信がないときは、第1または第2のネットワーク・コントローラ手段のもう一方をプライマリなデータ通信パスとして選択するものであってもよい。

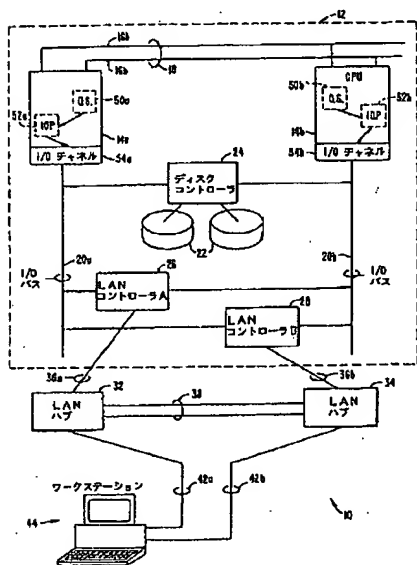
【0062】また、請求項4に記載の装置は、周辺装置およびこの周辺装置を前記入出力バス手段に接続する周辺コントローラ手段を含むこともできる。この周辺装置には、データ記憶手段が含まれる。

#### 【図面の簡単な説明】

【図1】この発明にしたがって構成されたネットワーク構成の概要を示すブロック図である。

【図2】コンピュータ・システムをネットワークに接続するネットワーク・コントローラの状態を追跡する図1のコンピュータ・システムにより保持される状態図である。

【図1】



【図3】図1に示すワークステーションをアクセス・デバイスに接続するために使用されるインタフェース装置の概要を示すブロック図である。

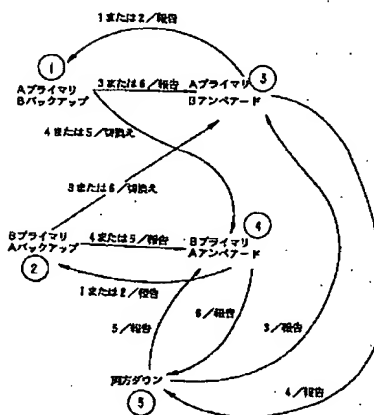
【図4】より複雑なコンピュータ・システムにおけるこの発明の使用を示す、より大規模なコンピュータ・システムのブロック図である。

#### 【符号の説明】

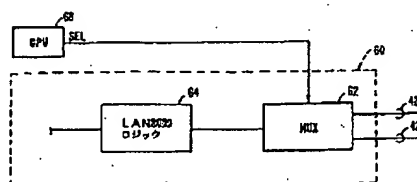
- 10 ネットワーク
- 12 コンピュータ・システム
- 14、14a、14b プロセッサ・ユニット
- 16、16a、16b バス
- 20、20a、20b 入出力バス
- 22 ディスク・ユニット
- 26、28 LANコントローラ
- 32、34 ハブ
- 38 リンク
- 42a、42b リンク
- 44 ワークステーション

20

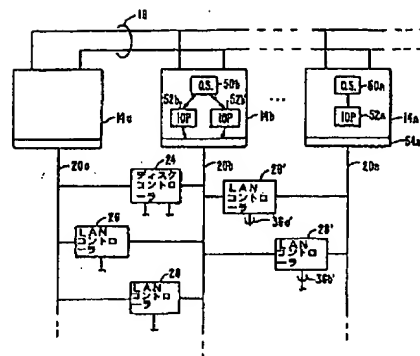
【図2】



【図3】



【図4】



フロントページの続き

(51)Int. Cl.<sup>6</sup>

G 0 6 F 15/16

識別記号

序内整理番号

4 7 0 A

F I

技術表示箇所